A Fast Algorithm for Mining Association Rules in Medical Image Data

Adepele Olukunle and Sylvanus Ehikioya,
Department of Computer Science,
University of Manitoba,
Winnipeg, MB, Canada, R3T 2N2.

{adepele, ehikioya}@cs.umanitoba.ca

Abstract

This paper presents a fast association rule mining algorithm which is suitable for medical image data sets. In particular, it assesses the feasibility of using association rule algorithms to extract hidden information from medical image data sets. In addition, we provide a flavour of our implementation environment. Finally, we show, with an example, how our proposed algorithm works to assess its suitability.

Keywords: Data Mining; Association Rules; Medical Images

1. INTRODUCTION

In the past decade, the use of image technology in medicine has changed enormously with advances in technology developments. Besides radiological images (X-rays), several other new digital imaging techniques such as Magnetic Resonance (MR), Ultasonograpy (US), Projectional Computed Radiography (CR), Computed Tomography (CT), and Position Emission Tomography (PET), have been widely used to assist physicians in decision making (diagnosis and prognosis). Information is extracted from these images on the basis of recognizing trends and patterns in previous examples. The emergence of these image technologies has resulted in a large collection of image data.

Images are rich in information content. Since the discovery of X-rays in 1895 [17], medical images have provided significant assistance in medical diagnosis. Medical images contain a wealth of hidden information that can be exploited by physicians in making reasoned decisions about a patient. However, extracting this relevant hidden information is a critical initial step to their use. This motivates the use of data mining techniques for efficient knowledge extraction.

Various data mining techniques (such neural networks, classification techniques, and cluster analysis) have been used to extract knowledge from medical data [2, 4]. These advances have been driven by the fact that huge data sets are continuously collected from patients, and at present most of the data are accessible electronically [3]. However, the applicability of these techniques within the medical image domain has been limited due to the complexity of the images, their use, and interpretation. Medical experts are prone to easily accept and apply transparent, explanative and evidential methods [4]. Thus, a need for a simple and explanative data mining technique suitable for the medical domain provides the motivation for this paper.

Association rule mining (ARM) [15] is a popular data mining technique, which aims at discovering strong interesting patterns (associations) between items in vast data sets. For instance, $X \Rightarrow Y$, read "X implies Y", is an association rule, which is interpreted as "if X occurs it is most likely that Y also will occur". The task of mining association rules is simply searching for items which occur frequently (large items), defined by a user-defined minimum frequency (minimum support), and then finding patterns in their occurrences. Various algorithms exist to efficiently search for and count large item sets. Hipp *et al.* [14] provide a detailed survey of ARM algorithms. A number of them have bottlenecks, which limit the size and nature of datasets they can efficiently mine. For instance, generating candidate items sets have always been a source of bottleneck in the implementation of Apriori–like ARM algorithms in large data sets. Another problem encountered is the I/O overhead, which occurs when the transaction database exceeds the size of main memory.

The Frequent-Pattern (FP) Growth algorithm [14, 22], however, is a standard ARM algorithm, which is efficient for mining large datasets with frequent patterns (i.e., a large number of "large" items). Its efficiency lies in the compact and complete way it represents the entire set of transactions and patterns in a tree-like form, which eliminates the need to generate candidate items sets. However, for extremely large datasets, the frequent–tree structure can also exceed the main memory, causing I/O overheads. We propose the *Partitioned Frequent-Pattern (FP) Growth algorithm*, which is based on partitioning the transaction database, and processing each of the partitions in parallel. This approach enables the efficient processing of large datasets with long patterns.

Various ARM techniques have been extensively applied to market basket, census and financial data mainly because of its simplicity and transparency [19]. Recently, ARM has been extended to the medical image domain.

Experiments in [13] indicate that ARM algorithms can be useful for medical image classification. The Apriori algorithm was applied to classify tumors extracted from a mammography database made up of 300 images. However, for a large database (which is often the case in medical image datasets), this approach is inadequate because it is limited by the large size of candidate items sets generated.

Our approach, however, involves using a variant of the FP-Growth algorithm, which in addition to having a compact data representation scheme, is amenable to parallel implementation and has an efficient input / output structure.

This study is significant for the following reasons:

- The proposed algorithm can assist physicians improve the accuracy and speed of decision-making based on medical images in medicine, thereby leading to dramatic improvements in health care.
- Experiences gained in applying the proposed algorithm to medical image data can provide the catalyst to promote [efficient] knowledge extraction from image data sets in other domains, e.g., geology, astronomy and biology.

This rest of this paper is organized as follows. Section 2 briefly reviews related literature on image mining. Section 3 describes the nature of medical images, the difficulties in applying ARM techniques to medical images, and key features of an ARM algorithm that will efficiently mine them. Section 4 describes the processes involved in image mining. Section 5 describes our proposed algorithm, our implementation environment and an example to illustrate its use. Finally, we conclude the paper in Section 6 with a brief summary and a roadmap for future directions.

2. RELATED WORK ON IMAGE MINING

Medical Images are generally mined using techniques such as clustering, classifier trees, or regression [18]. Lee [21] developed an intelligent decision making system for breast cancer diagnosis based on segmentation and classification using neural network. Inductive logic programming systems have also been used in *Ocular Fundis* image classification for glaucoma diagnosis [4]. Other machine learning methods, such as the Bayesian classifier, have also used in the diagnosis and prognosis of first cerebral paroxysm [7].

Furthermore, several classifiers have been used in the prognosis of the femoral neck fracture [2], in computer-aided diagnosis in chest radiography [16, 17] and also in the automated diagnosis and prognosis of breast cancer based on histological images [2].

Ordonez and Omiecinski [20] used the partitioned Apriori-based algorithm to mine simple geometrical shapes. Their choice of algorithm was influenced by the algorithms fast implementation on large data sets. Their work indicates that image mining is feasible using the ARM approach and simple rules can be obtained from a few simple objects. However, the candidate item sets generated limit the amount of memory occupied by discovered associations.

Association rules can be used in the medical domain in two ways. One way is to use ARM to validate the rules used by an expert system, e.g., ARM was used to validate the rules used by PERFEX [18], an expert system which aids heart disease diagnosis. The other way is to use ARM to discover new rules that relates causes to disease. For example, research in [11] focuses on the discovery of interesting associations between brain images and other clinical data. Both uses can be approached by associating the presence of certain items (symptoms) with some other items (disease feature) in the transaction (image), which can perhaps signify the presence of the disease. This use suits the binary nature of ARM itemsets, which are either present or not present. It is important that medical databases are mapped into transaction formats without losing their information. A transaction refers to the mapped record of all medical information belonging to a patient.

3. NATURE OF MEDICAL IMAGE DATA

Medical data sets are usually extremely large and are characterized by missing values, a temporal nature (time-ordered), long patterns (large number of attributes), and noise [4]. Hence in the application of mining tools for medical diagnosis and prediction, algorithms that can handle these irregularities are important. Medical image data sets are complex in nature [2]. A lot of pre-processing is required before medical images can be used in ARM algorithms.

To determine the necessary pre-processing required and provide some context for the appropriate ARM algorithms for medical data sets, we describe the features of medical images below.

- Medical image databases are heterogeneous, having different modalities, formats and resolutions [1], which require integration before they can be analyzed.
- Medical image data is usually multidimensional. These dimensions (two or three) must be represented in some form of attributes.
- Most images have a temporal component, especially if sets of images in the database represent a particular patient's history over time [10].
- Often image features are represented in qualitative values, which require discretization, with minimum loss in information before they can be mined.
- The value of medical image features are highly similar, thus, pattern detection must be sensitive. In this case, a low minimum support value will serve for sensitivity.
- The spatial correlation of objects in a medical image

must also be considered, as trends in location also add to the information to be mined.

- The multiple occurrences of objects are significant in medical images, and this must reflect in the representation of the region. It is important to note that typical ARM algorithms do not consider the multiple occurrences of an item within a transaction.

The task of manually extracting information (finding trends, patterns and anomalies) from images is a highly knowledge-based task, which involves similarity matching, and a good knowledge of the domain to which the image mining is to be applied.

Two key difficulties in applying association rules to medical image data sets are:

1. The fundamental mining algorithms assume that data sets contain simple numeric and symbolic entries, hence medical image have to be pre-processed in order to transform them for use in data mining techniques. This pre-processing requires the transformation of the internal representation of the images into a format suitable for the algorithms. These transformation operations are complex.
2. Medical images have a large number of attributes, therefore, the space of rules generated is usually very large and techniques for handling the data are generally I/O and memory intensive. Although a large number of rules are generated, not all of them may be significant. If the acceptance value of the minimum support is not well defined, the rules set may be too large. However, potentially useful rules could be excluded if the minimum support value is too high.

The complex nature of medical images indicates that to fully represent them for use in ARM, each image (transaction) will have a very large number of attributes (items). Thus, the data mining process will be I/O, processor, and memory intensive. An algorithm for mining medical image data should have some key features, which will make the mining process efficient. These desirable features are:

1. A fast algorithm, which will *efficiently search* the huge space for large item sets, with minimal consideration for small item sets.
2. It is important that the algorithm has an *efficient itemset counting strategy* due to the large number of images (transactions) involved. FP-growth-like algorithms count the database once and use a compact tree structure (usually much smaller than the original database) to represent the large items.
3. It is desired that we *scan the transaction database the minimum number of times possible (an efficient I/O structure).* Algorithms that implement efficient pruning of small itemsets, e.g., the FP-Growth algorithm, also achieve minimal scanning of the transaction database.
4. An algorithm that can represent *long patterns in a complete and compact way* is necessary. Completeness implies that it should never break the pattern of any transaction, while compactness implies reduction of irrelevant information. These properties are found in the FP-growth algorithm.
5. An algorithm, which is *scalable to parallel implementation*, is desired. Serial algorithms are constrained by the memory and processor power limitations of a single processor. For large databases, such as image data sets, the transaction database or the candidate item sets can be partitioned and run on several processors, which work independently. This approach can provide for faster implementations and the ability to handle higher computational complexity. The partition algorithm is scalable and amenable to parallel and distributed implementations.

These key features desired in any algorithm for mining medical image data sets resolve issues that are dependent on the complex nature of medical images.

4. THE IMAGE MINING PROCESS

Mining medical images involves many processes. A more detailed account of these processes is available in [17, 20]. The process to be used depends on the type and complexity of image to be mined. For instance, it is simpler to mine 2-dimentional x-rays as compared to 3-dimentional CT scans of the brain. However, some processes are fundamental to the task of medical image mining, regardless of the complexity of image. We briefly discuss these processes below.

Data Preprocessing: This stage consists of several processes. These processes include data normalization, data preparation, data transformation, data cleaning, and data formatting. Normalization techniques are required to integrate the different image formats to a common format. Data preparation alters images to present them in a format suitable for transformation techniques. Next, the image is transformed in order to obtain a compressed (lossless) representation of it, e.g., using wavelet transforms. Segmentation is done to identify regions of interest (ROI) for the mining task, usually achieved using classifier systems. The segmentation step finds corresponding regions within an image, since item sets are extremely large. Images are usually represented in pixels. A pixel is a dot of light on an evenly spaced rectangular grid, which has attributes such as color and texture. Collections of neighboring pixels with similar attributes make up a region. *Chu et al.* [10] proposed a temporal evolutionary data model, which enables the temporal and evolutionary descriptions of images, so that their image features and content can be modeled.

Feature Extraction: Images have a large number of features. It is important to identify and extract interesting features for a particular task in order to reduce the complexity of processing. These are attributes or portion of the image being analyzed that is most likely to give interesting rules for that problem. Not all the attributes of

an image are useful for knowledge extraction. This stage increases the overall efficiency of the system. Image processing algorithms are used, which automatically extract image attributes such as local color, global color, texture, and structure. Texture is the most useful description property of an image [6] and it specifies attributes, such as resolution, which can be used in image mining. An image can be adequately represented using the attributes of its features. The extraction of the features from an image can be done using a variety of image processing techniques [5, 8, 9 12]. We localize the extraction process to very small regions in order to ensure that we capture all areas.

Rule Generation: Since this is a highly knowledge-based domain, associated domain knowledge can be used to improve the data-mining task. Zrimaec [6] proposes a system, which combines global patient data with medical images. This data integration is an important concept because medical images are not self-contained, and are often used in conjunction with other patient data in the process of diagnosis. We expect association rules of two forms: (i) Image contents unrelated to spatial relationships, e.g., if an image has a texture X, it is likely to contain protrusion Y and (ii) Image contents related to spatial relationships, e.g., if X is between Y and Z it is likely there is a T beneath. A low minimum support and high minimum confidence is desirable, since few image data sets have high support [20].

Interpretation of patterns and knowledge extraction: Not all the interesting rules are medically important. To make our technique relevant, the rule presented must be significant and meaningful.

5. THE PARTITIONED FP-GROWTH ALGORITHM

Our proposed algorithm has the desired characteristics and can achieve faster processing (speedup), when compared to the sequential version of the algorithm. We present our proposed algorithm below.

Algorithm: Partitioned Frequent Pattern (FP) Growth

Divide Transaction Database into n partitions for n number of processors

for i = 1 to n **do begin**
  count 1-item sets in partition i
endfor

Read Min supp // User-defined minimum support for large item sets.

Sum local counts for each item X, in each partition to obtain Global count of item, S_x^G

If S_x^G > = Min supp // Prune all items that are not large
  Store large items in Header table, H in descending order of S_x^G
endif

for i = 1 to n **do begin**
  Construct FP-tree$_i$ in each Partition i
  Mine the FP-tree to generate FP-conditional pattern base for item x
  Merge all conditional pattern bases for each item X in partition P_i to form local pattern β_x^i
endfor

Merge local patterns to form global pattern, β_x^G for item x

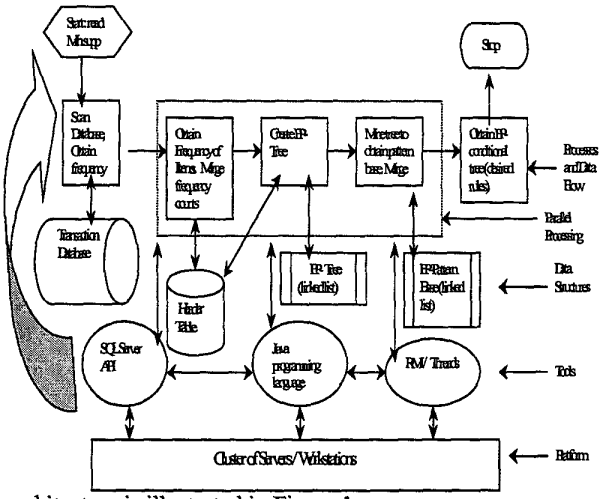
if support in global pattern. β_x^G < Minsupp
  Prune X
else
  Mine the global FP-Pattern base, β_x^G to obtain the global conditional FP_tree
endif

if *tree* $\beta_x^G \neq \phi$ then
  Mine tree, β^{k_G}
endif;

Mine association rules from global FP-tree;

end;

Implementation: The prototype platform consists of a cluster of Windows-based servers and workstations. Java and C programming serve as implementation languages. Communication between processors is via message passing. Each transaction in the database has a unique transaction id (TID). Every transaction consists of a TID and a list of items. The database has a maximum of five items per transaction. The transaction database and the header table are stored in an SQL server database. The FP-trees are implemented using tree-like structures (like linked lists) in which each element is linked to the proceeding and succeeding one. The implementation



architecture is illustrated in Figure 1.

Fig1: Implementation Environment.

Results: We obtain a global FP-conditional tree, a linked list of the large items, which shows the patterns between them. This pattern is interpreted directly to obtain the desired rules.

EXAMPLE

We illustrate how our algorithm works with an example. Given a sample database, T, with 40 transactions (T1 - T40) and four partitions (P1- P4). The database has a maximum of 5 items per transaction, representing features of interest extracted from an imaginary mammography database. The features of interest are:

- Large Object Size (where an object is a pixel in the segment of interest): L
- Low Noise Level: N
- High Contrast: C
- Coarse Texture (fractural-based): T
- Abnormal Average Grey Level: A

Let us assume it is known that an abnormal average grey level and a coarse texture range are present in all known abnormal cases. We need to study the associations between these particular features and some other selected features, e.g., high contrast, low noise and large object size, which will enable us imply the probability of an abnormal case from the presence of these features. An object with a particular feature value is said to have that item, otherwise it does not have it.

Our goal is to obtain associations between the various values of important features. It is interesting to see if the presence of a particular value in one feature tends to imply another value. The transactions in each partition are shown below. The size of each partition may not necessarily be the same.

Transaction	Items	Partition
T1	C, A	P1
T2	T, N, C	P1
T3	L, T	P1
T4	T, L, C,	P1
T5	L, N, C	P1
T6	T, C, L	P1
T7	T, A, C	P1
T8	A, L, T	P1
T9	C, N, L	P1
T10	T, C, L	P1
Transaction	**Items**	**Partition**
T1	A, C, N	P2
T2	T, N	P2
T3	L, C	P2
T4	N, L, C	P2
T5	L, C, N	P2
T6	C, L, N	P2
T7	N, L, C, T	P2
T8	T, C, L	P2
T9	C, T	P2
T10	L, A	P2
Transaction	**Items**	**Partition**
T1	L, N, C	P3
T2	A, C, N	P3
T3	A, C, N	P3
T4	C, N	P3
T5	L, C, T	P3
T6	C, N, L	P3
T7	T, L, N	P3
T8	L , T, N	P3
T9	A, L	P3
T10	T, C, L	P3
Transaction	**Items**	**Partition**
T1	L, T, N	P4
T2	A, T, L, N	P4
T3	L, C	P4
T4	L, C, T	P4
T5	T, C, N	P4
T6	N, C, T	P4
T7	L, C, N	P4
T8	T, A, L	P4
T9	N, L, C	P4
T10	T, N, A	P4

We assume a user defined minimum support (Minsupp) value of 12.

Algorithm Run

1. Partition T into 4 partitions P1 to P4
2. For each partition, P_i, count 1-itemsets.

P1

Item	L	N	C	T	A
Frequency	7	3	8	7	3

P2

Item	L	N	C	T	A
Frequency	7	6	8	4	2

P3

Item	L	N	C	T	A
Frequency	7	7	7	4	3

P4

Item	L	N	C	T	A
Frequency	7	7	6	6	3

3. Merge counts for items in all partitions.

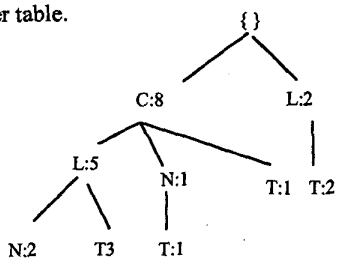
Item	L	N	C	T	A
Total Frequency	28	23	29	21	11

4. Prune items less than minimum support of 12. So, A is removed from the itemsets since its frequency is less than 12.
5. Rearrange frequency table in descending order of frequency to form header table.

HEADER TABLE

Item	C	L	N	T
Frequency	29	28	23	21

6. Construct the FP- tree for P1, using the order in the header table.



7. Similarly, construct FP-tree for P2, P3 and P4.
8. Mine FP-trees to obtain conditional pattern base for item branch in each partition.

Patterns for Tree 1:	Patterns for Tree2:
N2: C, L	T1: C, L
T3: C, L	T1: C, L, N
T1: C, N	N1: C
T1: C	T1: C
T2: L	T1: N

Patterns for Tree 3	Patterns for Tree 4
	T1: N
T1: C, L	T1: L
T2: L, N	T1: C, L
N2: C, L	T2: L, N
N3: C	T2: C, N
	N2: C, L

9. Merge local patterns for each item to obtain global conditional pattern base.

Partition	Item	Frequency Sums	Pattern Base
[1, 2, 3, 4]	T	(3+1+1+1=6)	C, L
[1,2]	T	(1+1=2)	C
[1,4]	T	(1+2=3)	C, N
[3]	T	(1)	C, L, N
[2,3]	T	(2+2 = 4)	L, N
[2,3]	N	(1+3 = 4)	C
[1,3]	N	(2+2+2=6)	C, L
[1,4]	T	(2 +1 =3)	L
[2,4]	T	(1+1 = 2)	N

Other intermediate patterns exist in the trees but ignored such patterns since we are only interested in the patterns that include T (our interesting feature. Mining the T pattern bases recursively, we obtain:

Partition	Item	Frequency Sums	Pattern base	Pattern
[1, 2, 3, 4]	T	(3+1+1+1 = 6)	C, L	T (6): C, T (6): L, T (6): CL
[1,2]	T	(1+1 = 2)	C	T (2): C
[2, 4]	T	(1+1=2)	N	T (2): N
[1,4]	T	(2+1 = 3)	L	T (3): L
[1,4]	T	(1+2 = 3)	C, N	T (3): C, T (3): N, T (3): C, N
[3]	T	(1)	C, L, N	T (1): C, T (1): L, T (1): N, T (1): C, L, T (1): C, N, T (1): C, L, N
[2,3]	T	(2+2 = 4)	L, N	T (4): L T (4): N, T (4): L, N

Adding the values for each item, we obtain

Item	Frequency Sums	Pattern
C	6+2+3+1 = 12	C ⇒ T
L	6+ 1+4 +3 = 14	L ⇒ T
L, N	4	L, N⇒ T
C, L	6+ 1 = 7	C, L ⇒ T
N	3+1+4 = 8	N ⇒ T
C, N	3+1 +2 = 6	C, N ⇒ T
C, L, N	1	C, L, N ⇒ T

With a support of 12, only the rules L →T and C → T are interesting. We can, thus, interpret our result to mean an object with a Large object size, L or with a high contrast, C, (or both), will most likely have a coarse texture T, which invariably classifies the object as abnormal.

6. CONCLUSIONS

Medical images contain a lot of information. It is desirable to mine them in order to use such knowledge for diagnosis support. ARM is a major data mining technique, which is simple and explanative. To apply an ARM technique to medical images, such a method must be able to handle huge data sets efficiently. The partitioned FP-Growth algorithm we proposed can mine medical image data sets efficiently.

There remains a need to apply the proposed algorithm to actual medical image data sets. This constitutes part of our future work. It will also be interesting to investigate the applicability of our proposed algorithm in other domains. Furthermore, a performance analysis of this algorithm is necessary to validate our conclusions. Optimizing the construction and mining processes of the FP-tree may further enhance the efficiency of the image mining process.

References

1. P. Cabena, P. Hadjinian, R. Stadler, J. Verhees and A. Zanasi. *Discovering Data Mining: From Concept to Implementation.* Prentice Hall, 07458,1997, pp. 12- 30

2. N. Lavac, E. Keraunou and B. Zupan. "Intelligent Data Analysis in Medicine and Pharmacology." *(IDAMAP-97),* Nagoya, Japonska, Kluwer Acedemic Publishers, 1997, pp 61-67.

3. W. Horn, "Artificial Intelligence in Medicine on its Way from Data-Intensive to Knowledge-Intensive", *Austrian Research Institute for Artificial Intelligence Technical Report TR-2001-01,* Vienna, Austria, Vol 23. No 1, 2001 pp. 5-12

4. N. Lavrac, "Data Mining in Medicine: Selected Techniques and Applications", *In: Proc. of the Second International Conference on The Practical Applications of Knowledge Discovery and Data Mining*, pp 11-31.

5. S. Startchik, *Geometric and Illumination Invariant Object Representation: Application to Content-based Image Retrieval.* Ph.D. Dissertation No. 3009, University of Geneva, Switzerland, July 1998.

6. T. Zrimaec, "A Medical Image Information System" *VISIM Workshop*, Utrecht, Netherlands, Oct. 2001, Paper 2.

7. G.D. Magoulas and A. Prentza, "Machine learning in Medical Applications", *Workshop on Machine Learning in Medical Applications* (ACAI-99), 1999, pp.53-58.

8. P. A. Devijer and J. Kitter, *Pattern Recognition: A Statistical Approach,* Prentice-Hall, 1982.

9. K. Woods, *Automated Image Analysis Techniques in Digital Mammography*, Ph.D. Thesis, Department of Computer Science and Engineering, University of South Florida, December 1994.

10. W. Chu, I. Leong, R. Taira, C. Breant, "A Temporal Evolutionary Object Oriented Data Model and its Query Language for Medical Image Management", *Proceedings of the 18th VLDB conference*, Vancouver, Canada, 1992, pp.53—64.

11. V. Megalooikonomou, J. For, L. Shen, F Makedon. Data Mining in Brain Imaging. *Statistical Methods in Medical Research* 2000, pp. 359-394.

12. O. R. Zaiane, *Resource and Knowledge Discovery from the Internet and Multimedia Repositories,* PhD. Thesis, School of Computing Science, Simon Fraser University, March 1999.

13. M. Antonie, O. Zaiane, A. Coman, "Application of Data Mining Techniques for Medical Image Classification" *In Proceedings of the Second International Workshop on Multimedia Data Mining,* with *ACM SIGKDD Conference (MDM/KDD'2001),* San Francisco, USA, August 26, 2001.

14. J. Hipp, V. Guntzer and G. Nakhaeizadeh, "Algorithms for Association Rule Mining. A General Survey and Comparison." *ACM SIGKDD* Volume 2, Issue 1, July 2000 pp. 58 –64.

15. R. Aggrawal and R. Srikant, "Fast Algorithms for Mining Association Rules" In *Proceeding of the 20th International Conference of Very Large Data Bases (VLDB),* Chile, 1994, pp 487- 499.

16. Ginneken, B. Romeny. Statistical Local Texture Analysis Applied to Computer- Aided Diagnosis in Chest Radiography Statistics of Shapes and Textures. *Copenhagen summer school in computer vision* September 4-8 2000

17. B. Ginnekan. *Computer Aided Diagnosis in Chest Radiography*. PhD thesis, Utrecht University, March, 2001 pp. 14-17.

18. C. Ordonez, C. Santana and L. Braal, "Discovering Interesting Association Rules in Medical Data", *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery,* 2000, pp. 78-85.

19. C. Aggrawal and P. Yu, "Mining Large Itemsets for Association Rules", *In Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.* Vol.2, No.1, 1998, Pg 23-31.

20. C. Ordonez and E. Omiecinski. "Discovering Association rules Based on Image Content", *In Proceedings of the IEEE Advances in Digital libraries Conference (ADL '99),* Pg. 38- 49.

21. K. Lee, "Intelligent Shape based Image Analysis", *Department of Computer Science, University of Iowa, 55:247* Project, 2001.

22. J.Han, J. Pei and Y. Yin, "Mining Frequent Patterns without Candidate Generation" *In Proceedings of SIGMOD-2000*, Dallas, May 2000,pp 1-12.